

### **REMARKS**

The present Amendment is in response to the Office Action mailed February 23, 2007. Claims 2-3 and 19-20 were previously cancelled. Claims 1, and 17-18 are currently amended. Claims 1, 4-18, and 21-34 remain pending in view of the above amendments and the following remarks.

Please note that the following remarks are not intended to be an exhaustive enumeration of the distinctions between any cited references and the claimed invention. Further the remarks herein or lack of remarks are not to be construed as an admission on the part of the Applicant regarding any assertion made by the Examiner or any characterization of the prior art set forth by the Examiner. Rather, the distinctions identified and discussed below are presented solely by way of example to illustrate some of the differences between the claimed invention and the cited references. Reconsideration of the application is respectfully requested in view of the above amendments to the claims and the following remarks. For the Examiner's convenience and reference, Applicant's remarks are presented in the order in which the corresponding issues were raised in the Office Action.

### **Terminal Disclaimer**

The terminal disclaimer filed on 11/19/06 was objected to as co-pending application 10/729,883 has not been identified and was included in the non-statutory obviousness-type double patenting rejection. A corrected terminal disclaimer is submitted herewith as requested by the Examiner.

### **Informalities**

The Examiner asks the Applicant to acknowledge certain informalities. The Examiner noted that claim 6 has been repeated twice in the prior response. In the prior response, claim 7 was inadvertently numbered as claim 6. Claim 7 has now been properly numbered as required by the Examiner. The Examiner also noted that the limitation "using linguistic information to extract" in claim 17 should have been underlined so as to indicate that it was a new limitation not present in prior versions of

the claim. This limitation has been underlined in this response to acknowledge the informality as required by the Examiner.

### **Rejection Under 35 U.S.C. §102**

The Office Action rejected claims 1, 4-18 and 21-34 under 35 U.S.C. § 102(b) as being anticipated by "Information Extraction: Beyond Document Retrieval" (*Gaizauskas*). Because *Gaizauskas* does not teach or suggest each and every element of the rejected claims, Applicants respectfully traverse this rejection in view of the amendments herein and the following remarks.

*Gaizauskas* describes information extraction "whose function is to extract information about a pre-specified set of entities, relations or events from natural language texts and to record this information in structured representations called templates." See abstract. The information extraction described by *Gaizauskas*, however, has substantial differences from the claimed invention and fails to teach or suggest several elements of the pending claims.

For example, *Gaizauskas* teaches the use of an information retrieval (IR) system to retrieve a corpus of articles. The result of the query is illustrated in Figure 1b of *Gaizauskas*. In contrast to the requirements of claim 1, however, the query returned by the IR system includes structured and unstructured data that are not independent.

In this context, independence is the ability for a piece of data to have meaning, or value, without regard to other data. A good example would be a person's drivers license number and their passport number. Both numbers identify the same person, but each number can exist on its own and be used for some purpose in isolation. Figure 1b of *Gaizauskas*, for example, illustrates a retrieved text that has structured data such as a document number (e.g., <DOCNO>) and a headline (e.g. <HL>). Neither the document number nor the headline make much sense (or have much value) without the document to which they are related. Thus, the structured data of Figure 1b is not independent from the unstructured data, i.e. the text. Further, the unstructured data of Figure 1b is not derived independently from the structured data.

Indeed, since these structured data describe the associated document, they comprise "metadata." Metadata are structured, encoded data that describe characteristics of related information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities. Where the information bearing entity described by the metadata consists of unstructured text (such as in Figure 1b), the metadata typically are considered to be "document metadata." Document metadata often contain the name of the person who created the file, the organization where the file originated, the date the file was created or modified, and other information about the document. *Gaizauskas* teaches only the use of metadata, and specifically, document metadata.

Claim 1, in contrast, has been amended to clarify that the unstructured data including the free text is derived independently of the structured data. For example, the specification discloses that free text can be added to a database when a customer calls a call center – notes from that conversation are often transcribed and added to the database as a text field. Thus, the text field is derived independently of the structured data. Examples of structured data that are then associated with that text field might include a customer ID, a product model or serial number for the product about which the customer, or the purchase history of the customer. Each of those three values, is derived independently from the text in the text field: a customer ID may originate with the Finance Department and may be used to identify any accounts receivable owed by that customer; a model number may originate in a system that retrieves user manuals and other product documentation; and the purchase history may originate with or be used by the Marketing Department to identify repeat customers for targeted mailings.

Note that these data are characteristically different from the kinds of structured data fields taught by *Gaizauskas*. *Gaizauskas* does not teach that the structured data fields are derived independently of the text. One way to recognize the distinction is to identify when the structured data are created and assigned to the unstructured content. In *Gaizauskas*, the document ID and the headline (or title) are typically created along with the text. In that sense, the structured and unstructured fields are concurrent with

each other. This would be typical of texts that could be thought of as articles or documents, i.e. the kind of text found in publications.

Claim 1, in contrast, clarifies that the unstructured data is derived independently of the structured data. In fact, the structured data may be created at any time – before or after the free text is stored in a text field. A customer ID, for example, is likely generated at the time of purchase and would be already in existence before a service call comes in and unstructured text notes are captured. Alternatively, the purchase history of the customer will change in the future when the customer buys another product, and the call notes from today may be referenced and analyzed after the purchase history has been updated. These structured data are typically not associated with publications but with operational information systems in businesses in which transactions are stored as records in a database.

In short, the independence of the structured data may lie in the purpose, time, place or process of the creation of the structured data, as well as distinctness in purpose, meaning or use of the data (i.e., semantic distinctness).

Extracting facts from both the unstructured data in a document and the document's own metadata, as taught by *Gaizauskas*, does not teach or suggest that the unstructured data is derived independently of the structured data. Thus, relating extracted facts from unstructured data to independent or semantically distinct structured data, which may have originated from a different time, place or source, or might have a purpose or meaning unrelated to the unstructured text data, is novel and is not taught by the cited art.

One advantage of the claimed invention is the ability to then extract relational facts from the free text (i.e. from the unstructured data) and then relate those facts back to independent structured data, i.e. data that do not describe the free text and that have meaning and utility independent of it. After extracting relational facts from the free text and producing a set of construed data from said unstructured data, claim 1 has been amended to clarify that each construed datum is then related to the semantically distinct structured data of the database record in which the free text was found.

In order to teach these requirements of claim 1, *Gaizauskas* would need to show that the information extraction taught by *Gaizauskas* is then related to other semantically distinct structured data rather than just document metadata as illustrated in Figure 1b. However, the requirement of relating each construed datum to such independent structured data of a database record as required by claim 1 is not taught or suggested.

Instead, *Gaizauskas* uses the structured data and unstructured data illustrated in Figure 1b to fill a predefined template as illustrated in Figure 1c. All of the fields in that template describe the contents of the text. There is no teaching or suggestion of relating the filled template back to data that are independent of the article. As taught by *Gaizauskas*, information extraction "may be seen as the activity of populating a structured information source (or database) from an unstructured, or free text, information source." See page 1.

*Gaizauskas* may suggest the activity of populating a structured information source, but this fails to teach or suggest relating construed datum back to independent, structured data of a database record from which the free text was drawn as is required by claim 1.

For at least these reasons, Applicants respectfully submit that claim 1 is patentable over the cited art. For at least the same reasons, independent claims 17 and 18 are also patentable over the cited art. For example, claim 17 has been amended to clarify that the unstructured data including the free text is added to the data records independently of the structured data and that the structured data reflects information captured separately from the unstructured data. Claim 17 later requires relating each construed datum to the structured data. As discussed above, these aspects of claim 17, among others, are not taught or suggested by *Gaizauskas*, which instead fills a predefined template and does not relate the results back to the original structured data. Claim 18 has similar amendments and is patentable for at least the same reasons. The pending dependent claims are therefore patentable for at least the same reasons.

### **Conclusion**

In view of the foregoing, Applicants believe the claims as amended are in allowable form. In the event that the Examiner finds remaining impediment to a prompt allowance of this application that may be clarified through a telephone interview, or which may be overcome by an Examiner's Amendment, the Examiner is requested to contact the undersigned attorney.

Dated this 23<sup>rd</sup> day of August, 2007.

Respectfully submitted,

/Carl T. Reed/ Reg. # 45454

CARL T. REED

Registration No. 45,454

Attorney for Applicant

Customer No. 022913

Telephone No. 801.533.9800

W:\10422\1096\MJV\0000001370V001.doc